(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平7－104780

(43)公開日　平成７年(1995)４月21日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　3/00	５３１　Ｊ	9379－5H		
	Ｚ	9379－5H		
	５６１　Ａ	9379－5H		

審査請求　有　　請求項の数５　ＯＬ　（全　８　頁）

(21)出願番号　特願平5－247835

(22)出願日　平成５年(1993)10月４日

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地

(72)発明者　山口　耕市
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内

(72)発明者　嵯峨山　茂樹
東京都保谷市中町５丁目５番10号

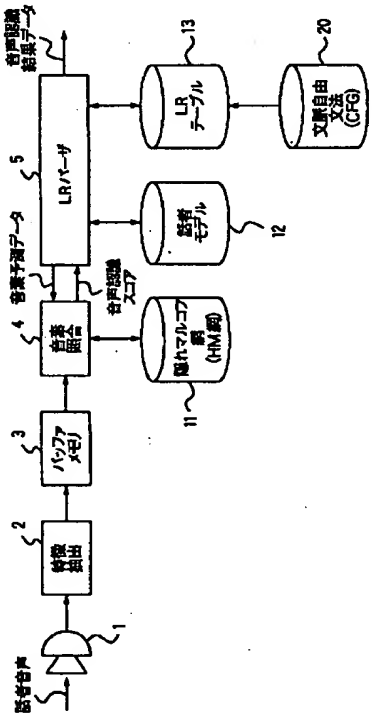
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　不特定話者連続音声認識方法

(57)【要約】

【目的】　従来例に比較して計算量を軽減することができ、しかも音声認識率を大幅に改善することができる不特定話者連続音声認識方法を提供する。

【構成】　複数ｍ人の話者に対応して複数ｍ個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された１人の話者の発声内容に基づいて発声内容と話者の２方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、それ以外のときは、尤度がしきい値未満となった枝に接続される１つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。



1

【特許請求の範囲】

【請求項１】　不特定話者の音声を連続的に認識する不特定話者連続音声認識方法であって、

複数ｍ人の話者に対応して複数ｍ個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された１人の話者の発声内容に基づいて発声内容と話者の２方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、

上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される１つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定することを特徴とする不特定話者連続音声認識方法。

【請求項２】　上記音声認識動作中又は完了後は、最大の尤度を有する話者を選出し、選出した話者を標準話者として話者モデルに対して話者適応することを特徴とする請求項１記載の不特定話者連続音声認識方法。

【請求項３】　上記音声認識動作中又は完了後は、所定の上位複数個の尤度を有する話者を選出し、選出した話者群を標準話者群として話者モデルに対して話者適応することを特徴とする請求項１記載の不特定話者連続音声認識方法。

【請求項４】　上記音声認識動作完了後に、最大の尤度を有する１人の話者の発声内容を選出し、選出した発声内容を教師信号として用いて話者モデルに対して話者適応することを特徴とする請求項１記載の不特定話者連続音声認識方法。

【請求項５】　請求項１記載の不特定話者連続音声認識方法において、

上記音声認識動作完了後に、最大の尤度を有する１人の話者の発声内容を選出し、選出した１人の話者を入力話者として判断することによって話者識別することを特徴とする話者識別方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、不特定話者の音声を連続に認識する不特定話者連続音声認識方法に関する。

【０００２】

【従来の技術及び発明が解決しようとする課題】従来の不特定話者音声認識システム（以下、第１の従来例という。）が、例えば、Madeleine Bates et al.：“Design and Performance of HARC, The BBN Spoken Language Understanding System”, Proc. ICSLP-92, pp. 241-244(1992年)に開示されている。この第１の従来例においては、「不特定話者音響モデル」が用いられ、多数の話者

2

の音声データを混合してしばしば学習に用いることが多いために、広がり過ぎた混合分布によって認識性能が劣化する恐れがある。また、ユーザは音声の登録なしに使えるという利点がある反面、認識しにくい話者に対して性能を改善することができないという欠点がある。

【０００３】また、英語の母音認識を行う方法（以下、第２の従来例という。）が、P. Niyogi and V. W. Zue: “Correlation Analysis of Vowels and their Application to Speech Recognition”, Proc. Eurospeech-91, pp. 1253-1256(1991年)に開示されている。この第２の従来例においては、母音の相関的な解析を音素認識に適用しているが、文法を用いて連続音声認識を行っていない。

【０００４】さらに、入力された話者音声に対して、１つの男子音響モデルと１つの女子音響モデルとを用いて音声認識処理を並列に実行し、入力音声の最後において最高の音声認識スコアを有した認識候補を認識結果とする方法（以下、第３の従来例という。）が、例えば、V. Abrash et al., “Connectionist Gender Adaptationin a Hybrid Neural Network/Hidden Markov Model Speech Recognition System”, Proc. ICSLP-92, pp. 911-914(1992年)において開示されている。この第３の従来例においては、音響モデルを２種類しか用いていないために、不特定話者音響モデルのような広がり過ぎた音響モデルによる認識性能の劣化が考えられる。また、複数の音響モデルを用いて音声認識処理を並列に実行する方法（以下、話者並列方法という。）を用いているために計算量が大きくなり、認識効率が比較的悪いという問題点があった。

【０００５】上記話者並列方法において、標準パターンを話者S_i毎に設定してそれぞれ音声認識とビームサーチとを並列に実行させ、入力音声の最後に最も高いスコアの仮説を認識結果とすることが考えられる。図３にこの話者並列方法を用いた従来の不特定話者認識モードにおける音素解析木を示す。図中、各々の枝に沿って記されたアルファベットは予測・照合された音素を表す。図３の音素解析木を用いて音声認識を実行する装置においては、各枝毎に複数人分の話者の尤度を予め記憶しており、複数人分のモデルに対して最尤度を計算するためには、最後の音声まですなわち音素解析木の最右側の端部まで並列にすべての複数人分のモデルに対する計算を実行する必要があるため、計算量が大きくなり、認識効率が悪くなるという問題点があった。

【０００６】本発明の目的は以上の問題点を解決し、従来例に比較して計算量を軽減することができ、しかも音声認識率を大幅に改善することができる不特定話者連続音声認識方法を提供することにある。

【０００７】

【課題を解決するための手段】本発明者は、入力される発声音声は不特定話者の発声音声であっても、その話者

は発声を通じて同一であることに着目して以下に示す本発明を発明した。本発明に係る請求項１記載の不特定話者連続音声認識方法は、不特定話者の音声を連続的に認識する不特定話者連続音声認識方法であって、複数ｍ人の話者に対応して複数ｍ個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された１人の話者の発声内容に基づいて発声内容と話者の２方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される１つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定することを特徴とする。

【０００８】また、請求項２記載の不特定話者連続音声認識方法は、請求項１記載の不特定話者連続音声認識方法において、上記音声認識動作中又は完了後は、最大の尤度を有する話者を選出し、選出した話者を標準話者として話者モデルに対して話者適応することを特徴とする。さらに、請求項３記載の不特定話者連続音声認識方法は、請求項１記載の不特定話者連続音声認識方法において、上記音声認識動作中又は完了後は、所定の上位複数個の尤度を有する話者を選出し、選出した話者群を標準話者群として話者モデルに対して話者適応することを特徴とする。また、請求項４記載の不特定話者連続音声認識方法は、請求項１記載の不特定話者連続音声認識方法において、上記音声認識動作完了後に、最大の尤度を有する１人の話者の発声内容を選出し、選出した発声内容を教師信号として用いて話者モデルに対して話者適応することを特徴とする。さらに、請求項５記載の話者識別方法は、請求項１記載の不特定話者連続音声認識方法、上記音声認識動作完了後に、最大の尤度を有する１人の話者の発声内容を選出し、選出した１人の話者を入力話者として判断することによって話者識別することを特徴とする。

【０００９】

【実施例】以下、図面を参照して本発明に係る実施例について説明する。本実施例の不特定話者連続音声認識方法は、図２にその一例を示す音素解析木上のビームサーチする方法を用いて、発声内容と話者の２方向をサーチの対象とし、尤度が最大である発声内容と話者とを同時

$$P(w, S_i \mid y) = P(y \mid w, S_i)\, P(w)\, P(S_i) / P(y)$$

ここで、Ｐ（S_i）は入力話者が第ｉ番目の話者である先験確率である。本実施例において、確率Ｐ（ｗ）は、好ましくは０．１に設定する。また、不特定話者音声認識タスクを対象としているので、Ｐ（S_i）はすべての

に決定して音声認識することを特徴とする。音声認識動作中に計算された尤度が所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される図２の右方向の１つ又は複数の枝を枝刈りする。そして、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。

【００１０】本実施例の不特定話者連続音声認識方法について説明するために、まず、定式化を行う。１つの発話が多数の話者の声で構成されることは現実にはほとんどあり得ない。すなわち、音声認識システムの対象が不特定話者であっても、１つの文や単語列などの中ではすべての音素は同一の話者によって発声されるという制約がある。この原理的な制約を「話者一貫性原理」と呼ぶ。

【００１１】まず、話者一貫性原理の数学的定式化を行なう。ｗを単語列$w = w_1, w_2, \cdots, w_n$とおく。音響パラメータの時系列ｙが与えられたならば、音声認識処理は次の数１を満足する最大の尤度を有する単語列（最尤単語列）ｗａを見付けることである。ここで、「ｗａ」の「ａ」は最大尤度を示す添字である。

【数１】

$$P(wa \mid y) = \max_{w} P(w \mid y)$$

ここで、右辺のｍａｘは単語列ｗに関して確率Ｐ（ｗ｜ｙ）の最大のものを表わす。Ｐ（ｗ｜ｙ）は音響パラメータの時系列ｙが与えられたときに単語列ｗが見つかる確率であり、Ｐ（ｗａ｜ｙ）は、単語列ｗに関する確率Ｐ（ｗ｜ｙ）の中で最大（ｍａｘ）の確率を有する最尤単語列ｗａの確率である。

【００１２】ここで、１つの文や単語列などの中ではすべての音素は同一の話者によって発声されるという上記話者一貫性原理を、数１に適用すると次の数２を得る。

【数２】

$$P(wa, Sa \mid y) = \max_{w,\, S_i} P(w, S_i \mid y)$$

【００１３】ここで、右辺のｍａｘは単語列ｗとｉ番目の話者S_iに関する確率Ｐ（ｗ，S_i｜ｙ）の最大のものを表わす。S_iは第ｉ番目の話者（ｉは１，２，…，ｍのいずれかである。）を表す。上記数２の右辺をベイズの定理を用いて書き換えることによって次の数３を得る。

【数３】

話者（ｉ＝１，２，…，ｍ）に対して等確率とする。

【００１４】上記数２と数３から、音声認識処理の目的は積Ｐ（ｗ）Ｐ（S_i）Ｐ（ｙ｜ｗ，S_i）を最大にする単語列ｗａおよび話者Ｓａを見付けることに相当するの

5　　6

で、次の数４を得る。

【数４】

$$P(wa)P(Sa)P(y|wa,Sa)=\max_{w,\ S_i}\{P(w)P(S_i)P(y|w,S_i)\}$$

【００１５】ここで、右辺のｍａｘは、単語列ｗとｉ番目の話者S_iに関する｛Ｐ（ｗ）Ｐ（S_i）Ｐ（ｙ｜ｗ，S_i）｝の最大のものを表わす。Ｐ（ｙ｜ｗ，S_i）は単語列ｗがある話者（ｉ番目の話者）S_iによって制限されていることを意味する。すなわち、本方法は発話内容の単語列ｗに加え、話者｛S_i｝も探索の対象とする。認識動作完了とともに、選出された最大の尤度を有する話者が、以後の話者適応で使うのに適した標準話者Ｓａとして選出される。

【００１６】上記話者一貫性原理の別の定式化も可能であって、最終的な尤度はすべての話者を対象とすることによって、次の数５を得る。

【数５】

$$P(wa)P(y|wa)=\max_{i=1}^{m}\{P(w)\Sigma P(S_i)P(y|w,S_i)\}$$

【００１７】ここで、右辺のΣはｉ＝１からｍまでの代数和である。数５は、すべての話者による寄与を考慮に入れているということを意味する。数５の場合においては、ビームサーチのときに枝刈りされて出て来ないパスが出てくるので、最大の尤度の単語列のみならず、別の話者方向を加えて別のパスも加算してサーチする。この場合、最大の尤度を有する話者Ｓｍは別途求める必要がある。

【００１８】以上に述べた話者バージング方法を、図１に示すＳＳＳ（Successive State Splitting：逐次状態分割法）－ＬＲ（left-to-right rightmost型）不特定話者連続音声認識装置に適用する。この装置は、メモリ１１に格納された隠れマルコフ網（以下、ＨＭ網という。）と呼ばれる音素環境依存型の効率のよいＨＭＭの表現形式を用いている。また、上記ＳＳＳにおいては、音素の特徴空間上に割り当てられた確率的定常信号源（状態）の間の確率的な遷移により音声パラメータの時間的な推移を表現した確率モデルに対して、尤度最大化の基準に基づいて個々の状態をコンテキスト方向又は時間方向へ分割するという操作を繰り返すことによって、モデルの精密化を逐次的に実行する。

【００１９】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して音素照合部４に入力される。

【００２０】音素照合部４に接続されるＨＭ網メモリ１１内のＨＭ網は、各状態をノードとする複数のネットワークとして表され、各状態はそれぞれ以下の情報を有する。

（ａ）状態番号
（ｂ）受理可能なコンテキストクラス
（ｃ）先行状態、及び後続状態のリスト
（ｄ）出力確率密度分布のパラメータ
（ｅ）自己遷移確率及び後続状態への遷移確率

【００２１】なお、本実施例において、話者バージングのためのＨＭ網は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合ＨＭ網を変換して作成する。ここで、出力確率密度関数は３４次元の対角共分散行列をもつ混合ガウス分布であり、各分布はある特定の話者のサンプルを用いて学習されている。

【００２２】音素照合部４は、音素コンテキスト依存型ＬＲパーザ（以下、ＬＲパーザという。）５からの音素照合要求に応じて音素照合処理を実行する。このときに、ＬＲパーザ５からは、音素照合区間及び照合対象音素とその前後の音素から成る音素コンテキスト情報が渡される。音素照合部４は、受け取った音素コンテキスト情報に基づいてそのようなコンテキストを受理することができるＨＭ網上の状態を、先行状態リストと後続状態リストの制約内で連結することによって、１つのモデルが選択される。そして、このモデルを用いて音素照合区間内のデータに対する尤度が計算され、この尤度の値が音素照合スコアとしてＬＲパーザ５に返される。このときに用いられるモデルは、隠れマルコフモデル（以下、ＨＭＭという。）と等価であるために、尤度の計算には通常のＨＭＭで用いられている前向きパスアルゴリズムをそのまま使用する。

【００２３】文脈自由文法データベースメモリ２０内の所定の文脈自由文法（ＣＦＧ）を公知の通り自動的に変換してＬＲテーブルを作成してＬＲテーブルメモリ１３に格納される。ＬＲパーザ５は、例えば音素継続時間長モデルを含む話者モデルメモリ１２と上記ＬＲテーブルとを参照して、入力された音素予測データについて左から右方向に、後戻りなしに処理する。構文的にあいまいさがある場合は、スタックを分割してすべての候補の解析が平行して処理される。ＬＲパーザ５は、ＬＲテーブルメモリ１３内のＬＲテーブルから次にくる音素を予測して音素予測データを音素照合部４に出力する。これに応答して、音素照合部４は、その音素に対応するＨＭ網メモリ１１内の情報を参照して照合し、その尤度を音声認識スコアとしてＬＲパーザ５に戻し、順次音素を連接

していくことにより、連続音声の認識を行っている。複数の音素が予測された場合は、これらすべての存在をチェックし、ビームサーチの方法により、部分的な音声認識の尤度の高い部分木を残すという枝刈りを行って高速処理を実現する。入力された話者音声の最後まで処理した後、詳細後述するように、全体の尤度が最大のもの又は所定の上位複数個のものを認識結果データ又は結果候補データとして出力する。

【００２４】本実施例の連続音声認識装置においては、音素解析木上のビームサーチを採用している。図２は話者パージング認識モードにおける音素解析木を示し、各々の枝に沿って記されたアルファベットは予測・照合された音素を表わす。ビームサーチによってある話者の仮説がすべて枝刈りされてしまうことがあるため、数５は近似的にしか用いることができない。従って、本実施例の実際の装置では数４を採用する。まず最初に複数ｍ人の話者S_i（ｉ＝１，２，…，ｍ）に対応してｍ個の仮説が存在する。その後、各仮説は話者毎にそれぞれ音素に同期にして成長し、ビームサーチにより枝刈りされる。すなわち、認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残すが、一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される図２の右方向の１つ又は複数の枝を枝刈りして、これ以降、認識候補として残さず、尤度の計算を行わない。そして、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者を同時に音声認識結果として決定する。本実施例においては、音素解析木は音素に同期して成長されているが、これに限らず、時間軸のフレームに同期して成長させてもよい。

【００２５】従って、本実施例においては、発声内容と話者の２方向を同時にサーチの対象とし、最大の尤度の発声内容と話者を同時に決定することを特徴とし、上述のように、ビームサーチによって発声内容と話者の２方向の仮説を枝刈りする。

【００２６】また、音声認識動作中又は完了後は、最大の尤度の話者を選出し、選出された話者を標準話者として話者適応してもよい。話者適応は、具体的には、入力話者の発声音声の少量の音声データを用いて以下のように行われる。標準話者の特徴ベクトルを入力話者の特徴ベクトル空間へ写像する移動ベクトルをＨＭＭの学習により求め、この写像を用いて話者適応を行う。この方法は、この写像の連続性と滑らかさを仮定することにより、少量の音声データによる話者適応の高精度化を実現している。すなわち。学習により移動ベクトルが得られなかった特徴ベクトルについては、近傍の特徴ベクトルをもちいて内挿する。また、データ不足に対しては、得られた移動ベクトルに平滑化を施す。さらに、尤度が所定の上位複数個の話者を選出し、音声認識動作中又は完了後は、選出された話者群を標準話者群として話者適応してもよい。

【００２７】本実施例の方法は、認識結果の最尤単語列ｗａを利用することができる点から、本方法は、言語制約を取り入れた教師信号なしの話者適応装置に適用することができる。また、最大の尤度の１人の話者の発声内容を選出し、認識動作完了後に、選出された発声内容を教師信号として用いて話者適応してもよい。さらに、最大の尤度の１人の話者の発声内容を選出し、認識動作完了後に、選出された１人の話者を入力話者として判断することによって話者識別してもよく、これにより話者識別装置を構成してもよい。

【００２８】本発明者は、本実施例の図１に示す装置を用いて文節単位でシミュレーションを行い、本発明に係る話者パージング方法と、従来技術の不特定話者法、及び第３の従来例の話者並列方法との認識性能を比較した。

【００２９】まず、当該シミュレーションの条件は以下の通りである。評価話者は１２名（男性５名、女性７名）であり、評価データは３４５文節からなる「国際会議予約タスク」を用いた。従って、全データ数は４，１４０文節となる。文脈自由文法のルール数は２，８１３であり、音素パープレキシティは３．３であった。ＨＭ網の状態数は２００であり、混合数は２０であった。ビーム幅は最大１，２００に設定した。なお、話者並列方法ではビーム幅２，４００でも行なっており、このときの１話者あたりのビーム幅はそれぞれ６０と１２０に相当する。

【００３０】次いで、ＨＭ網は以下のようにして作成した。まず初期モデルとして１７０名（男性８５名、女性８５名）分の特定話者ＨＭ網を作成した。次に、この１７０名分のＨＭ網からクラスタリングによって２０個のＨＭ網（男性１１名、女性９名）を選出した。最後に各クラスタに属するメンバーの話者のサンプルを用いて、ＶＦＳ法によって再学習することで話者クラスタＨＭ網を作成し、それらを話者混合して不特定話者ＨＭ網とした。

【００３１】上記３つの方法による認識結果を表１に示す。この結果では、不特定話者を、話者パージング法はわずかに上回った程度である。本発明で用いた話者一貫性原理は、対象とする話者の種類が多い場合に有効であると考えられ、また話者パージング方法を用いた方が有意に上回っている評価話者も存在することから、今後、本方式の本質的な有効性が明らかになると期待できる。一方、話者並列方法は話者毎の仮説に対してビームサーチを行なうため、枝刈りの効率が悪く、無駄な話者の仮説が生き残っていることが多い。従って、ビーム幅を２，４００に設定してもなお、本発明の話者パージング方法と従来技術の不特定話者に及ばない。

【００３２】

【表１】

9　　10

文節認識率（%）

方法	不特定話者	話者パージング	話者並列	話者並列
ビーム幅	１２００	１２００	１２００	２４００
１位	８３．３	８３．９	６２．３	７６．１
１位～５位	９４．９	９５．２	７０．５	８６．１

【００３３】上記シミュレーションにおいては、文節単位で音声認識を行っているが、これに限らず、文単位又は複数の文単位で音声認識を行ってもよい。

【００３４】以上説明したように、不特定話者の発声であっても、話者は発声を通して同一である点に着眼した不特定話者連続音声認識方法である「話者パージング」方法を発明した。本発明者による上記ＳＳＳ－ＬＲ連続音声認識装置上で不特定話者音声認識シミュレーションを行ない、従来技術の不特定話者法との認識性能を比較した。今回のような小規模の実験においては不特定話者法の認識率をわずかに上回った程度であったが、本発明の方法は、将来、対象とする話者のバラエティが広い場合にその効果を発揮し、音声認識率を大幅に改善することができると考えられる。

【００３５】本発明に係る本実施例の不特定話者連続音声認識方法は、以下の特有の利点を有する。

（ａ）順位の低い仮説しかもたない話者は枝刈りされ、その時点から以後その話者の尤度は計算する必要がなくなり、HMMのフレーム尤度計算量が削減でき、これによって高速に処理することができる。例えば、２０個の話者クラスタをもつ音響モデルのとき、標準話者として１個の話者クラスタ（話者クラスタとは、複数の話者を含む１つのグループをいう。）を採用したならば尤度計算量は１／２０になる。

（ｂ）話者適応機能を用いることにより、話者の音響モデルを入力話者に効率よく適応させるとともに、不特定話者モードでは認識しにくい話者に効果的に対処させることができる。さらに、話者適応のための教師信号として認識動作完了後に選出した尤度最大の発声内容を用いることにより、「教師なし話者適応」が実現することができる。

（ｃ）従来の不特定話者音声認識システムは、「不特定話者音響モデル」が用いられ、多数の話者の音声データを混合して学習に用いたために、広がり過ぎた混合分布によって認識性能の劣化を有していた。これに対して、本実施例では、多数の話者の音声データを混合して学習する必要がないために、認識性能の劣化を回避することができ、これによって、高い認識性能を得ることができる。

【００３６】

【発明の効果】以上詳述したように本発明によれば、不特定話者の音声を連続的に認識する不特定話者連続音声認識方法であって、複数ｍ人の話者に対応して複数ｍ個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された１人の話者の発声内容に基づいて発声内容と話者の２方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される１つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。従って、本発明は以下の特有の効果を有する。

（ａ）順位の低い仮説しかもたない話者は枝刈りされ、その時点から以後その話者の尤度は計算する必要がなくなり、HMMのフレーム尤度計算量が削減できる。これによって、高速に処理することができる。

（ｂ）従来の不特定話者音声認識システムは、「不特定話者音響モデル」が用いられ、多数の話者の音声データを混合して学習に用いたために、広がり過ぎた混合分布によって認識性能の劣化を有していた。これに対して、本発明では、多数の話者の音声データを混合して学習する必要がないために、認識性能の劣化を回避することができ、これによって、高い認識性能を得ることができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施例である不特定話者音声認識装置のブロック図である。

【図２】　本実施例における話者パージング認識モードにおける音素解析木を示す図である。

【図３】　従来例における不特定話者認識モードにおける音素解析木を示す図である。

【符号の説明】

１…マイクロホン、
２…特徴抽出部、
３…バッファメモリ、
４…音素照合部、
５…ＬＲパーザ、
１１…隠れマルコフ網メモリ、
１２…話者モデルメモリ、

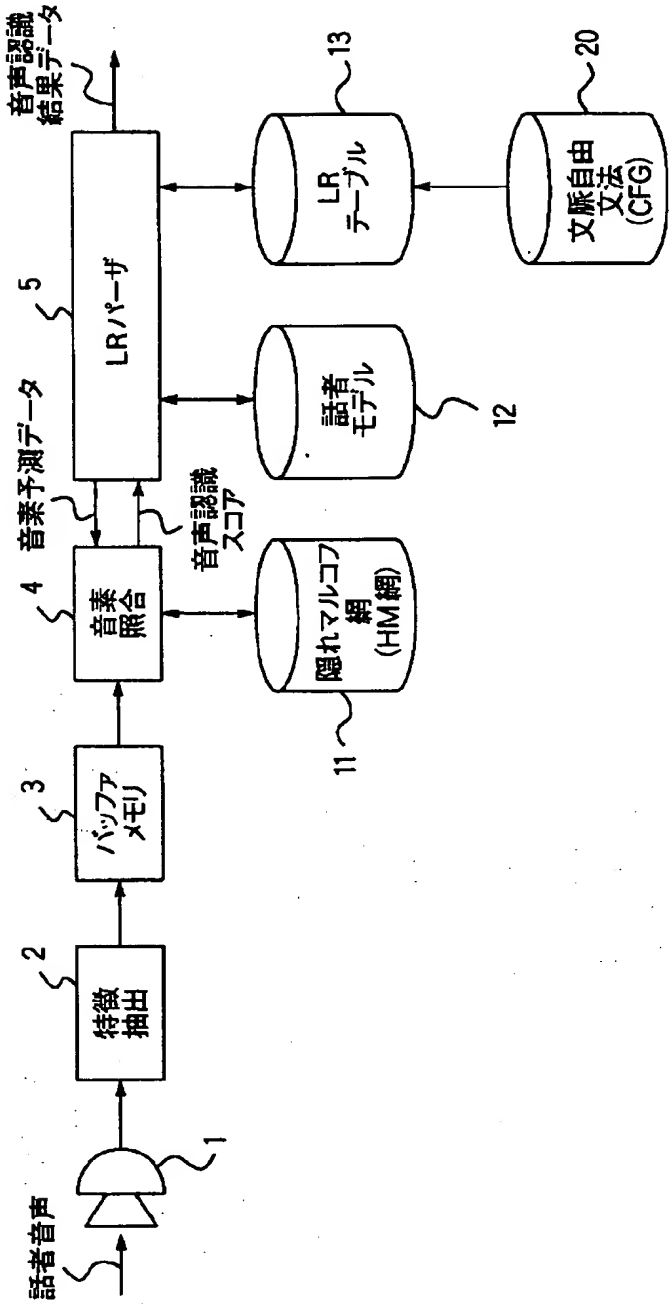
11

１３…ＬＲテーブルメモリ、

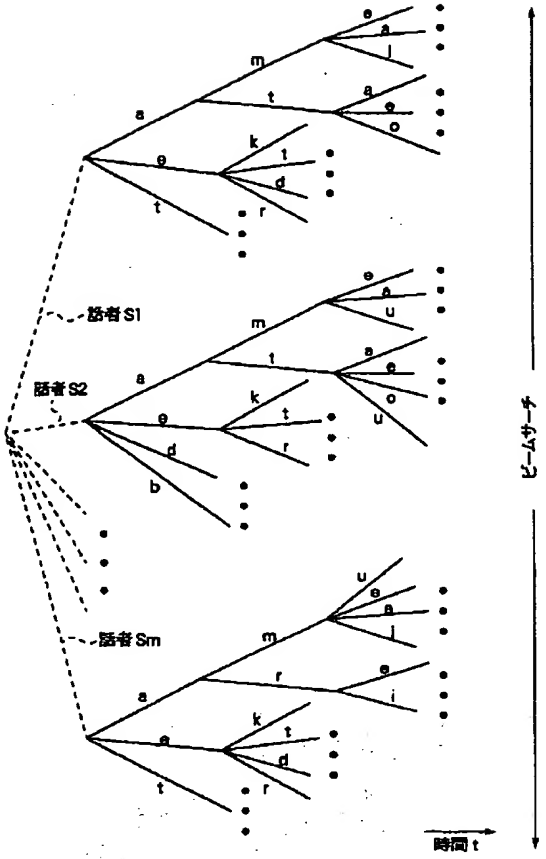
12

２０…文脈自由文法データベースメモリ。

【図１】



【図２】



【図３】